

COURT OPINIONS SHOW PATH TOWARD MORE DEFENSIBLE DOCUMENT REVIEW¹

BY KELLY R. YOUNG, ESQ.

A trio of recent opinions² from two different U.S. district courts, including one issued in late May by U.S. Magistrate Judge Paul Grimm, suggest how discovery of electronically stored information (“ESI”) will be conducted in the very near future. Of more immediate import, however, the three opinions provide useful guidance as to how discovery of ESI should be conducted today in order to successfully defend against challenges and thereby avoid incurring the massive costs and legal risks that come with losing such challenges. All three opinions guide litigators to a powerful new basis for defending reviews and productions of ESI: a scientifically based search methodology. Judge Grimm’s recent opinion goes one step further, advising parties and courts more specifically to look to The Sedona Conference Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery,³ as well as to the federal government’s TREC Legal Track,⁴ when making reasonableness and defensibility determinations with respect to the search methodologies underlying the production of ESI.

Reasonableness of Process

On May 29, 2008, Magistrate Judge Paul Grimm issued a memorandum opinion in copyright infringement case *Victor Stanley, Inc., v. Creative Pipe, Inc.*,⁵ rejecting the defendant’s post-production claim of attorney client privilege over documents that the defendant had produced to the plaintiff. Of interest to producing parties generally, Judge Grimm provides explicit guidance as to how ESI should be reviewed and produced, and, as a consequence, elucidates how productions can be successfully defended.

Considering the issue of waiver, Judge Grimm explains,

The intermediate test [for waiver of privilege by inadvertent production] requires the court to balance the following factors to determine whether inadvertent production of attorney-client privileged materials waives the privilege: (1) the reasonableness of the precautions taken to prevent inadvertent disclosure; (2) the number of inadvertent disclosures; (3) the extent of the disclosures; (4) any delay in measures taken to rectify the disclosure; and (5) overriding interests in justice. [*Victor Stanley* at 17-18]

It is the first factor—the reasonableness of the precautions taken by the party to prevent the inadvertent disclosure of otherwise privileged material—that holds the key for parties looking to stay ahead of the curve.

1 The analysis contained in this document is intended to be informational only and should not be construed as legal advice.

2 *United States v. O’Keefe*, 537 F. Supp. 2d 14 (D.D.C. 2008); *Equity Analytics, LLC, v. Lundin*, 248 F.R.D. 331 (D.D.C. 2008); and *Victor Stanley, Inc., v. Creative Pipe, Inc.*, 2008 WL 2221841 (D.Md.). Page references in this document correspond to the page numbers in the hyperlinked PDFs of these opinions.

3 8 Sedona Conf. J. (2007)

4 See: http://www.thesedonaconference.org/content/miscFiles/TREC_OPEN_Letter.pdf

5 2008 WL 2221841 (D.Md.)

The defendant utilized a search protocol based on 70 keywords that was agreed upon by the plaintiffs for purposes of responsiveness to plaintiffs' request for relevant material. The protocol was then used partially for privilege review. Judge Grimm determined that this search and retrieval protocol fell woefully short of establishing reasonableness:

While keyword searches have long been recognized as appropriate and helpful for ESI search and retrieval, there are well-known limitations and risks associated with them, and proper selection and implementation obviously involves technical, if not scientific knowledge. [*Id.* at 19]

For the Court to determine whether the search methodology was reasonably designed to prevent the inadvertent disclosure of privileged information, Judge Grimm explained that the defendant would have to have articulated:

1. Which keywords were chosen and how were they used to search the document population;
2. The rationale for selecting those keywords;
3. The qualifications of the individuals selecting such keywords to design effective searches; and
4. Whether and to what degree the results of the search were measured for reliability and quality. [*Id.* at 18]

The Court offered additional guidance for parties seeking to demonstrate the reliability and reasonableness of a particular search and retrieval methodology, pointing specifically to The Sedona Conference:

Selection of the appropriate search and information retrieval technique requires careful advance planning by persons qualified to design effective search methodology. The implementation of the methodology selected should be tested for quality assurance; and the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented. In this regard, compliance with the *Sedona Conference Best Practices* for use of search and information retrieval will go a long way towards convincing the court that the method chosen was reasonable and reliable ... [*Id.* at 26]

Without information regarding the steps a producing party actually took, why and how they took those steps, and the effectiveness of such steps in identifying the material of interest, how can one even begin to evaluate the reasonableness of the effort? One cannot. After all, as Judge Grimm reminds us, “[c]ommon sense suggests that even a properly designed and executed keyword search may prove to be over-inclusive or under-inclusive ...” [*Id.* at 11], a fact that is well known to every first-year law student who has searched for a helpful law review article or published case. So a rich set of keywords is not in and of itself evidence of reasonableness, especially without measurement of the search’s over- and under-inclusiveness.

Statisticians have long understood these false positives (over-inclusive results) and false negatives (under-inclusive results), and information retrieval science is largely about measuring them and controlling them. Recognizing this problem is, as Judge Grimm writes, a matter of

common sense. Such measurement and control, however, are far from common practice in electronic discovery. For those facing proverbial mountains of ESI, that situation is improving thanks to prudent applications of linguistic and statistical expertise.

Information retrieval science is making the measurement and control of false positives and false negatives readily available to litigants engaged in electronic discovery. And statistical science is making it realistic by providing tools that are cost-effective. As the Court states, “The only prudent way to test the reliability of the keyword search is to perform some appropriate sampling of the documents determined to be privileged and those determined not to be in order to arrive at a comfort level that the categories are neither over-inclusive nor under-inclusive.” [*Id.* at 11-12] Unfortunately for the defendants in *Victor Stanley*, they submitted no evidence of having conducted any such sampling.

Experts Are Needed

Some may accuse Judge Grimm of burying the lead by placing in his footnote 10 a great deal of information that is worthy of analysis all by itself. In that footnote, Judge Grimm refers to the two other recent decisions mentioned at the outset of this paper: *U.S. v. O’Keefe*⁶ and *Equity Analytics, LLC, v. Lundin*.⁷ Combined with *Victor Stanley* and some basic knowledge of the rules of expert evidence, *O’Keefe* and *Equity Analytics* guide discovery practitioners through the otherwise baffling world of search methodology, statistical sampling, expert qualifications, and measurement of results.

Less than four months ago in *O’Keefe*, U.S. Magistrate Judge John Facciola ruled that discovery of ESI was not exempt from the rules governing scientific and other expert evidence. A criminal case alleging bribery of a U.S. consular official, *O’Keefe* addressed a defendant’s challenge of the government’s production of ESI on the grounds that the government’s search terms were inadequate.

Judge Facciola rejected the defendant’s challenge as lacking sufficient legal basis. In particular, he noted that the challenge lacked any basis in the rules of expert evidence:

This topic [of search term efficacy] is clearly beyond the ken of a layman and requires that any such conclusion be based on evidence that, for example, meets the criteria of Rule 702 of the Federal Rules of Evidence [(“FRE”)]. Accordingly, if defendants are going to contend that the search terms used by the government were insufficient, they will have to specifically so contend in a motion to compel and their contention must be based on evidence that meets the requirements of [FRE] 702. [*O’Keefe* at 17]

In other words, to challenge the search methodology of a document production, a party must make a claim based on *expert* opinion, the admission of which is governed by FRE 702.

Significantly, Judge Facciola observed that making such a challenge is not simply arguing a point of law, much less negotiating more forcefully with opposing counsel. Instead, it is introducing expert evidence about the methodology of the search, i.e., the science of information

6 537 F. Supp. 2d 14 (D.D.C. 2008)

7 248 F.R.D. 331 (D.D.C. 2008)

retrieval. That may surprise many of us who work in electronic discovery day in and day out and are accustomed to negotiating keyword lists with contingent layers of Boolean expressions. But the notion that this work is best suited for true experts should not surprise any of us given the obvious complexities of searching gigabytes of data created in varied business contexts by multiple sources communicating in their unique professional dialects.

Taken together, *Victor Stanley, O'Keefe, and Equity Analytics* show the way to a well-grounded production. It is also clear how these productions will be defended as courts come to understand what Judge Facciola stated unequivocally: information retrieval is a science and requires legitimate experts. More eloquently, Judge Facciola explained:

Whether search terms or “keywords” will yield the information sought is a complicated question involving the interplay, at least, of the sciences of computer technology, statistics and linguistics.... Given this complexity, for lawyers and judges to dare opine that a certain search term or terms would be more likely to produce information than the terms that were used is truly to go where angels fear to tread. [*Id.* at 17]

How then do we put the guidance of these jurists into practice? Fortunately, the basis for the defense of such a production was articulated 15 years ago by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993). In *Daubert* the Court declared that truly scientific evidence had four basic criteria:

1. Falsifiability. The theory in question is testable. That is, one must be able to imagine a test for which some conceivable result would disprove the theory.
2. Peer Review. The theory or technique has been published and peer reviewed since that is how claims and suppositions become part of our common, validated scientific understanding.
3. Testing. The technique has been tested so that its error rate is known and so that it is subject to standards governing its operation.
4. Scientific Acceptance. The theory or technique is generally accepted by a *relevant* scientific community, not just any community and not just any so-called scientist.

Consequently, a party responding to a request for discovery should utilize measurable and actually measured techniques that rely on published, peer-reviewed information retrieval science. While attorneys choosing keyword search terms based on instinct or creativity may have sufficed in the past, the clock is fast running out on that approach.

Is discovery of ESI no longer governed by the Federal Rules of Civil Procedure and instead suddenly a creature of the Federal Rules of Evidence? Of course not. Clearly, the Federal Rules of Civil Procedure do not explicitly mandate the use of information retrieval experts in the discharge of discovery obligations. And FRE 702 formally governs only the admissibility of expert opinion testimony. So why should e-discovery practitioners suddenly consider themselves bound by *Daubert*? “Bound” may be too strong a term, but Judge Grimm makes the case for turning to FRE 702 and *Daubert*:

The goal of [FRE] 702 is to set standards to determine whether information is “helpful” to those who must make factual determinations involving disputed areas of science,

technology or other specialized information. The rule is one of common sense, and reason—opinions regarding specialized, scientific or technical matters are not “helpful” unless someone with proper qualifications and adequate supporting facts provided such an opinion after following reliable methodology. That these common sense criteria are found in the rules of evidence does not render them off-limits for consideration during discovery. It is not unusual for pretrial factual determinations in civil cases to look to the [FRE] for assistance in resolving fact disputes. [*Victor Stanley* at 22, fn. 10]

In other words, while the rules of civil procedure may not formally hold parties to the standards of FRE 702 and *Daubert*, courts are wise to look there when considering productions of ESI. And where courts look for guidance, parties should look for strategy. Failing to do so is missing an opportunity at the very least.

Judge Grimm then neatly sums it up:

The message to be taken from *O’Keefe, Equity Analytics*, and this opinion is that when parties decide to use a particular ESI search and retrieval methodology, they need to be aware of literature describing the strengths and weaknesses of various methodologies, such as *The Sedona Conference Best Practices, supra, n.9*, and select the one that they believe is most appropriate for its intended task. Should their selection be challenged by their adversary, and the court be called upon to make a ruling, then they should expect to support their position with affidavits or other equivalent information from persons with the requisite qualifications and experience, based on sufficient facts or data and using reliable principles on methodology. [*Id.* at 22, fn. 10]

What of Costs?

Do these opinions suggest that litigants must pay even greater sums to meet their already expensive discovery obligations? Are parties now expected to retain and pay for linguists, statisticians, and computer scientists in addition to their outside counsel? Again, Judge Grimm anticipates the concern and allays it:

For those understandably concerned about keeping discovery costs within reasonable bounds, it is worth repeating that the cost-benefit balancing factors of Fed. R. Civ. P. 26(b)(2)(C) apply to all aspects of discovery, and parties worried about the cost of employing properly designed search and information retrieval methods have an incentive to keep the costs of this phase of discovery as low as possible, including attempting to confer with their opposing party in an effort to identify a mutually agreeable search and retrieval method. This minimizes cost because if the method is approved, there will be no dispute resolving its sufficiency, and doing it right the first time is always cheaper than doing it over if ordered to do so by the court. [*Id.* at 22-23, fn. 10]

More importantly, the proper application of human expertise and technology offers the possibility of reducing discovery costs while achieving more accurate results, i.e., fewer false positives and fewer false negatives for fewer dollars. That may well require litigants to change their approach to discovery in significant ways, but those changes will bring improvements on many fronts.

Toward a Scientifically Valid Approach

The *Victor Stanley* opinion offers additional specific guidance that practitioners should heed if they want to remain on the leading edge in their field:

In addition, there is room for optimism that as search and information retrieval methodologies are studied and tested, this will result in identifying those that are most effective and least expensive to employ for a variety of ESI discovery tasks. Such a study has been underway since 2006, when the National Institute of Standards and Technology (NIST), an agency within the U.S. Department of Commerce, embarked on a cooperative endeavor with the Department of Defense to evaluate the effectiveness of a variety of search methodologies. This project, known as the Text Retrieval Conference (TREC), evolved into the Trec LegalTrack, a research effort aimed at studying the e-discovery review process to evaluate the effectiveness of a wide array of search methodologies. This evaluative process is open to participation by academics, law firms, corporate counsel and companies providing ESI discovery services. See: <http://trec-legal.umiacs.umd.edu>. The next test will occur in the summer of 2008. The goal of the project is to create industry best practices for use in electronic discovery. This project can be expected to identify both cost effective and reliable search and information retrieval methodologies and best practice recommendations, which, if adhered to, certainly would support an argument that the party employing them performed a reasonable ESI search, whether for privilege review or other purposes. [*Id.* at 23, fn. 10]

So the carrot: the promise of finding a better, more cost-effective approach than attorneys making up lists of keywords and unleashing them on document populations.

And the stick? Sanctions, quite possibly, and the prospect of a negative inference issued from the bench or other untoward result for counsel and client. While these three opinions do not directly address first-party discovery obligations, we can see the hand writing on the wall. If *Daubert*-compliant alternatives yield better results and cost the same or even less than alternative approaches, is there an argument for not deploying them? And for outside counsel seeking greater assurance that they are in fact fulfilling their professional responsibilities while protecting the best interest of their client, these opinions offer persuasive guidelines and comfort in the form of suggested best practices.

Whether a party is searching gigabytes of email or entire server farms of corporate records, a scientifically valid approach is called for simply because of the size and complexity of the challenge. Moreover, a scientifically valid approach will surely be demanded as parties and courts come to fully understand these decisions. The sooner litigants jump onto this discovery train, the less likely they are to be run over by it.

Mr. Kelly Young, Esq., is an attorney in Washington, DC, and practice director at H5. He can be reached at kyoung@h5.com.